On the use of an optimal database for radar forest observation

Laetitia Thirion-Lefèvre, Sandór Bilicz, Marc Lambert, Szabolcs Gyimothy

On the use of an “optimal database” for radar forest observation.

Laetitia Thirion-Lefèvre
SONDRA/Supélec at Gif-sur-Yvette, France

Sándor Bilicz, Marc Lambert
Département de Recherche en Électromagnétisme,
Laboratoire des Signaux et Systèmes (CNRS-SUPELEC-Univ Paris Sud), Gif-sur-Yvette, France

Szabolcs Gyimóthy
Budapest University of Technology and Economics, Egry J. u. 18, H-1521 Budapest, Hungary

Abstract

In this paper, we apply a fast method using an optimal database to estimate the impact of some radar sensor characteristics on three quantities (called “outputs”): the polarimetric scattering coefficient, the interferometric height and the total attenuation of the forested area under study. We consider two characteristics: the radar frequency and the incidence angle – called the “inputs” of the model. A variation domain for both of them is defined, leading to the so-called input space. Then, simulations are performed sequentially, controlled by a sophisticated strategy in order to build an “optimal database”. The latter is a set of corresponding input-output data pairs (samples), so that the output data are spread out as well as possible. In other words, the “output space” is filled by the samples of the optimal database completely and uniformly. The benefit of such an optimal database is multiple. First, the time-consuming numerical simulations can be replaced by a fast interpolation based on the samples stored in the database. Second, the structure of the optimal database reflects meta-information about the studied problem, helping us to see and understand better the behavior of the physics behind our model. This work might be a possible first step to reveal relationships between the observable radar characteristics and e.g., the true mean height of the forest.

1 Introduction

In a previous study, we have tried to relate the interferometric heights and the attenuation to the true mean height of the forest [1]. When the scattering mechanisms are restricted, this relation can be derived quite easily. For instance, we have illustrated that the true height is correctly estimated by the sum of the interferometric height and the penetration depth, provided that the waves do not penetrate below the crown. Indeed, in this case, we can assume that the direct scattering mechanism constitutes the main response (See **Figure 1**).

However, at low frequency, this condition is not satisfied and consequently the double bounces become more important. In this case, we have to face a non-linear problem. The interferometric height depends on the relative weight of the different scattering mechanisms, whereas the attenuation is derived using a trihedral corner reflector which *selects* the direct scattering mechanism. In this case, it is no more straightforward to relate the interferometric heights and the attenuation to the true mean height of the forest. To overcome this difficulty, the idea is to study the correlation between the interferometric height and the attenuation for different penetration depths. This can be achieved by considering wide ranges of frequency and incidence angle. To do so, we propose to use an adaptive sampling strategy which explores the studied problem step-by-step via the sequential evaluation of the model.

Such adaptive sampling methods are used in electromagnetic nondestructive evaluation (see, e.g., [2, 3]). In that case, the aim is to generate a *database* as a kind of discrete representation of the problem. A database consists of corresponding pairs of input parameter values - output data. By fitting an appropriate interpolator to these stored samples, approximate solutions of the underlying problem are available practically in real-time. However, the performance of such approximations strongly depends on the choice of the samples stored in the database – this is why adaptive sampling techniques are needed.

Though the main goal of generating adaptive databases is to obtain good approximations rapidly, the structure of such database also provides some meta-information about the problem to which it is adapted. In this paper, this secondary benefit is in the focus: how this meta-information can facilitate the analysis of hidden relations between the interferometric height and the attenuation in function of the radar frequency and incidence angle.

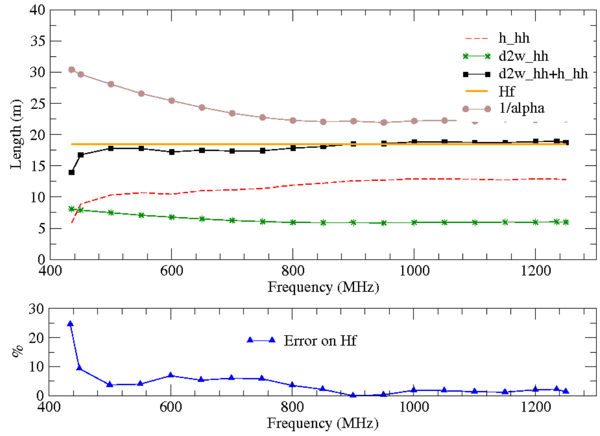

Figure 1: Top: Frequency variation of the interferometric phase center height h_{hh} (broken lines), the two-way penetration depth $d2w_{hh}$ (solid lines with stars) and the sum of these last two quantities (solid lines with squares).The mean height H_f of the forest is indicated (bold solid lines) as well as the frequency variation of the inverse of the attenuation (solid lines with circles). The 2-layers forest under study is $18.5\,\mathrm{m}$ high (trunks layer is $11\,\mathrm{m}$ high), radar frequency varies from $435\,\mathrm{MHz}$ to $1250\,\mathrm{MHz}$ and the incidence angle is about 58 degrees. The ambiguity height is constant and around $47\,\mathrm{m}$. Bottom: Frequency variation of the relative error between the true mean height H_f and the estimated height $d2w_{hh} + h_{hh}$.

2 The adaptive sampling strategy

Let us imagine a radar configuration characterized by k parameters, collected into a vector $\mathbf{x} = [x_1\,,\ x_2\,,\ \ldots\ ,\ x_k]^\mathsf{T}$, which is referred as the *input* of the model. The set of all conceivable inputs is the input space $\mathcal{X}$. Let $\mathbf{y}$ be the vector of the l *output* variables. The output space $\mathcal{Y}$ consists of all conceivable $\mathbf{y}$ outputs. There is no restriction for k and l. Let $||\cdot||_y$ be an appropriate norm on $\mathcal{Y}$, thus, a distance is defined on $\mathcal{Y}$. The connection between the input $\mathbf{x}$ and output $\mathbf{y}$ is represented by the so-called forward operator $\mathcal{F}$: $\mathbf{y} = \mathcal{F}\{\mathbf{x}\}$, which is evaluated by using a numerical simulator of the modeled scattering problem.

We use an incremental loop for the adaptive sampling [4]. The flowchart is presented in **Figure 2**. The process starts by a small number of initial samples, then, further samples are added (and $\mathcal{F}$ is evaluated) one-by-one according to a strategy till a stopping criterion is passed and the final database is built. The strategy tries to spread out the output samples in the output space $\mathcal{Y}$, i.e., to fill the output space evenly. The rule of the sample insertion is a kind of maximin method. The next $(n+1)$-th sample is added according to

$$\mathbf{x}_{n+1} = \arg\max_{\mathbf{x}\in\mathcal{X}} \left[\min_{i=1,\ldots,n} ||\mathcal{F}\{\mathbf{x}\} - \mathbf{y}_i||_y \right], \qquad (1)$$

i.e., the next output sample is aimed to be placed as far as possible from all other already added output samples. Note that in the implementation not the exact optimization problem (1) is solved, but it is recasted in a reduced form by using kriging prediction. This is needed due to the high computational cost of the model $\mathcal{F}$. Kriging provides a stochastic framework for function approximation (we approximate the functions $||\mathcal{F}\{\mathbf{x}\} - \mathbf{y}_i||_y$) and by now it has become a classical tool in several domains (for instance see [5] for a substantial overview). Kriging and its role in the sampling strategy is not discussed in this paper, see [4] for details. The incremental loop is running till a stopping criterion (the number of cycles in our case) is met.

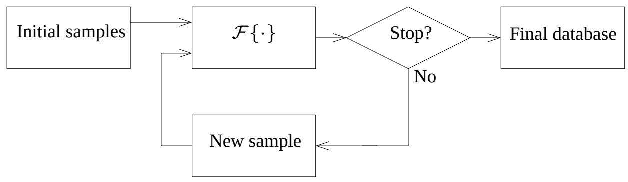

Figure 2: Flowchart of the adaptive sampling strategy.

3 Application to radar forest observations

The forest is a maritime pine tree forest, called Nezer forest, and is located in the south part of France. The ground truth data have been provided by INRA in the frame of the PYLA'2004 campaign, conducted by ONERA. Two different stands are available for study: stand 1, which is 52 years old and $26\,\mathrm{m}$ high and stand 2, which is 26 years old and $18.5\,\mathrm{m}$ high. These two stands are described accurately in [6] (the needles are not taken into account). For the radar configuration illustrated in **Figure 3**, we consider that the plane flies at an altitude of $3.5\,\mathrm{km}$. For the interferometric study, the baseline components may vary, but the ambiguity height remains constant around $100\,\mathrm{m}$.

Our study focuses on three quantities: the polarimetric scattering coefficient σ_{qp} where p, q refer to the polarization of the incident and scattered wave, which is either vertical or horizontal, the interferometric height h_{qp} and the total attenuation α_{pp} (α_{vv} or α_{hh}) of the forested area under study. Our long term purpose is to retrieve the mean height of the forest under observation and to relate it to its mean attenuation. As a first step, we study here the behavior of σ_{qp}, h_{qp} and α_{pp} with the radar frequency f and the incidence angle θ_i. Each of these quantities has to be ideally characterized by its statistical distribution. In this paper, we have considered the mean of these quantities. The standard deviation can be neglected for σ_{qp} and h_{qp}, but not for α_{pp}. However, it has not be taken into account in this study.

We define two variation domains for the input variables, leading to the input space $\mathcal{X}$: $0.3\,\mathrm{GHz} \le f \le 2\,\mathrm{GHz}$ and $40\,^\circ \le \theta_i \le 70\,^\circ$.

The output quantites (σ_{qp}, h_{qp} and α_{pp}) are treated separately, i.e., three different databases are generated. In all cases, the final number of samples in the database is 288. For all the three quantities and for two specific cases (at 435 MHz and 1.2 GHz and around $60\,°$), we obtained a good agreement between measurements and simulations [6]. One output sample is a vector of $[\sigma_{vv}, \sigma_{hh}, \sigma_{hv}]$, $[h_{vv}, h_{hh}, h_{hv}]$ or $[\alpha_{vv}, \alpha_{hh}]$, respectively. The output space $\mathcal{Y}$ is embedded in 3-dimensional (3D) space in the two first cases and it is 2D in the third case. The distance in $\mathcal{Y}$ is defined by the classical Euclidean norm in all cases.

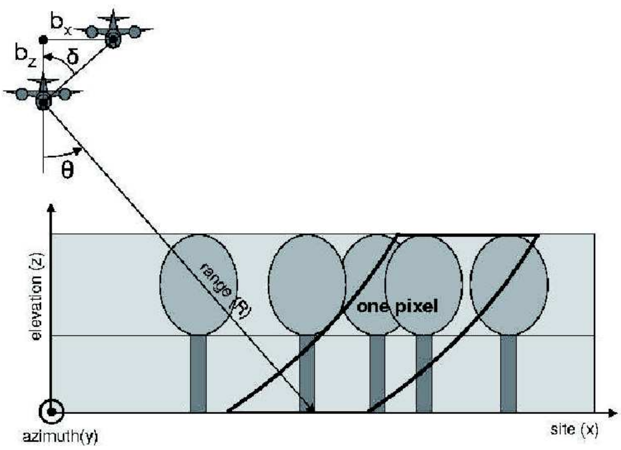

Figure 3: Radar configuration (side view).

The optimal database of the polarimetric scattering coefficients is presented in **Figure 4**. One can see that the output space (a 2D surface in a 3D space) is more-or-less evenly filled by samples, whereas in the input space much more samples are concentrated in the region of lower frequecies. The reason of this is that σ is more sensitive to the frequency in this region than in the domain of higher frequencies. If we look closer at the number of samples needed to fulfill the parts with frequencies below and above 600 MHz, we note that we need 10 times more data at the lower frequencies. This result is probably related to the structure of the forest, that may be described as a first layer of oriented scatterers (trunks) and then a second layer consisting in quasi-random volume (the crown). These two substructures may not require the same effort to be characterized as in the first case, we have to accurately described the scattering by an oriented element, whereas we know that few data are needed to represent correctly the scattering by a random volume. As we are rather in the lower part or in the upper one depending on the frequency, then the number of required data will vary. The backscattering coefficients vary in $[-17.5, -7.9]$ dB at VV polarization, in $[-16.9 - 3.2]$ dB at HH polarization and in $[-27.0, -16.1]$ dB at cross-polarization. Considering that the needles are not taken into account, these results are coherent.

In **Figure 5**, the database of the interferometric height h_{qp} is presented. Similar properties can be noticed as in the case of σ. If we divide the frequency domain as previously, we obtain that more than 3 times more measurements are needed in the low frequency part. Concerning now the values of the interferometric heights (not plotted), they vary in $[5.4, 18.4]$ m at VV polarization, in $[3, 18.1]$ m at HH polarization, and in $[10.9, 18.5]$ m at cross-polarization.

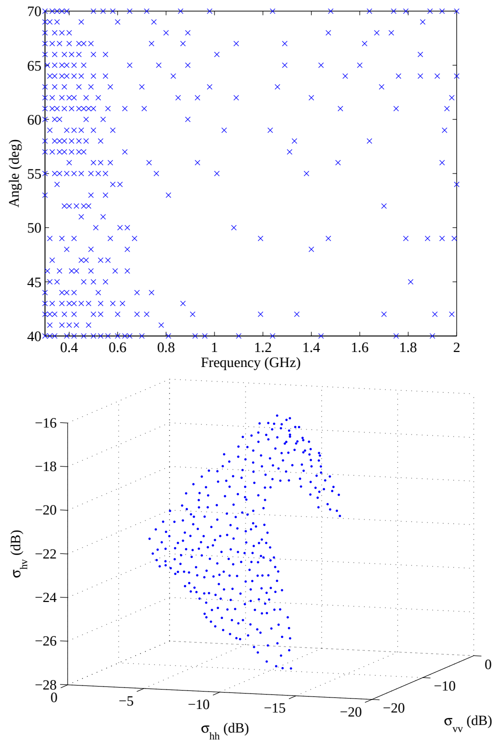

Figure 4: Optimal database of the polarimetric scattering coefficients σ_{qp}. Top: input space, bottom: output space.

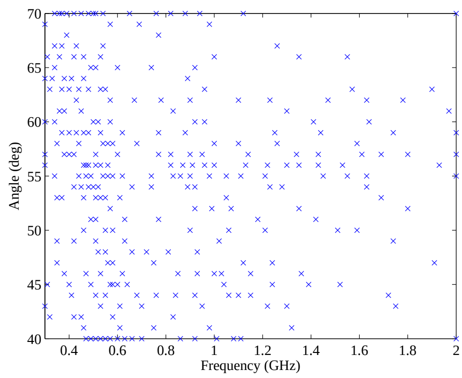

Figure 5: Optimal database of the interferometric height h_{qp}. (Only the input space.)

The total attenuation α_{pp} is easier to visualize as it has only two components in the output space. This is why we have generated a so-called regular database as well, in wich the input samples are spaced on a regular grid. Both databases are shown in **Figure 6**. It is easily seen that such a regular input sampling leads to quite distorted output sampling which is to be avoided, whereas our sampling strategy yields the optimal database. Concerning the value of the attenuation, there again the levels we obtain are coherent in comparison with our prevous studies [1]. However, this results have to be considered carefully as the distribution of the attenuation is very spread, leading to a standard deviation which is of the same order than the mean of α_{pp}. If we study separately what happens below and above 1 GHz, we clearly see that more points are required to describe the low frequency behaviour.

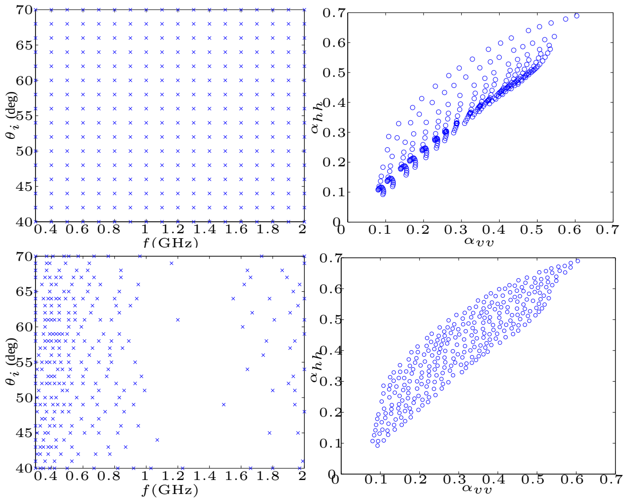

Figure 6: Regular (top) and optimal (bottom) database of the total attenuation α_{pp}. Left: input space, right: output space.

4 Conclusions

The authors have developed and applied an adaptive sampling method usually dedicated to speed up simulations. As expected, the time consumption has been greatly reduced, and from now on, we know that for this forest and for a frequency and an incidence angle included in the input space, we can simulate fast and easily either the polarimetric backscattering coefficients, the polarimetric interferometric heights or the polarimetric attenuation coefficients. This is interesting in order to fix some of the radar characteristics to study a specific area. In addition, the structure of the input space leading to a regularly sampled output space gives information about what kind of measurements are required to characterize the area. Indeed, we have observed that until about 600 MHz, we need to consider different incident angles (more than 27 angles per 100 MHz for the interferometric heights and more than 40 angles per 100 MHz for the backscattering coefficients). This is not the case above 600 MHz where only some incident angles are required (10 times less for the backscattering coefficients, more than 3 times less for the interferometric heights). This result seems logical regarding the sructure of the forest. At high frequencies, the crown of the forest mainly respond and if we can assume that it may be approximated as a random volume, consequently few incidence angles are required to characterize this reponse. On the contrary, at low frequencies, the waves penetrate until the trunks' level, and we need more incidence angles to describe the polarimetric response of these quasi-vertical elements. This result can also illustrate the relevancy of some measurements. Indeed at high frequency, we may expect that the the measurements give a quite good estimation of the forest's behaviour. At low frequency, the behaviour seems so versatile that we cannot expect to derive conclusions using only few measurements. For further studies, we may apply this technique to other forested areas in order to check our conclusions. Then, it might be interesting to define the number of required measurements depending on the accuracy we would like to obtain on the backscattering coefficients or on the interferometric heights. Finally, we would like to study the correlation between the different output spaces we obtain in order to state if we can derive total forest height using the attenuation and the interferometric heights, as we did at high frequencies in [6].

References

[1] L. Thirion-Lefevre and E. Colin-Koeniguer, "Investigating attenuation, scattering phase center, and total height using simulated interferometric sar images of forested areas," *IEEE TGRS*, vol. 45, no. 10, pp. 3172–3179, October 2007.

[2] J. Pávó and Sz. Gyimóthy, "Adaptive inversion database for electromagnetic nondestructive evaluation," *NDT&E International*, vol. 40, pp. 192–202, 2007.

[3] Sz. Gyimóthy, I. Kiss, and Pávó J., "Adaptive sampling technique based on moving meshes for building data-equidistant inversion databases for NDT," *International Journal of Applied Electromagnetics and Mechanics*, vol. 30., pp. 309–319, 2009.

[4] S. Bilicz, M. Lambert, E. Vazquez, and Sz. Gyimóthy, "A new database generation method combining maximin method and kriging prediction for eddy-current testing," in *Electromagnetic Non–Destructive Evaluation (XIII)*, Studies in Appl. Electromagn. and Mech. IOS Press, Amsterdam, 2010, 8 pages, **accepted**.

[5] J.P. Chiles and P. Delfiner, *Geostatistics, Modeling Spatial Uncertainty*, Wiley, 1999.

[6] L. Thirion, E. Colin, and C. Dahon, "Capabilities of a forest coherent scattering model applied to radiometry, interferometry, and polarimetry at p- and l-band," *IEEE TGRS*, vol. 44, no. 4, pp. 849–862, April 2006.